

ACADEMIE DES SCIENCES DE L'URSS

**L. Korolev, S. Rasoumovski,
G. Zélenkevitch**

**LES EXPERIMENTS DE LA TRADUCTION
AUTOMATIQUE DE L'ANGLAIS EN RUSSE
A L'AIDE DE LA CALCULATRICE БЭСМ**

MOSCOU — 1956

Approved For Release 2008/04/25 : CIA-RDP80T00246A004400020003-8

ACADEMIE DES SCIENCES DE L'URSS

L. Korolev, S. Rasoumovski,
G. Zélenkevitch

LES EXPERIMENTS
DE LA TRADUCTION AUTOMATIQUE
DE L'ANGLAIS EN RUSSE A L'AIDE
DE LA CALCULATRICE БЭСМ

MOSCOU — 1956

Approved For Release 2008/04/25 : CIA-RDP80T00246A004400020003-8

Le travail de la traduction automatique de l'anglais en russe a été commencé par l'Institut de Mécanique Exacte et de Technique de Calcul et par l'Institut d'Information Scientifique de l'Académie des Sciences de l'U.R.S.S. Les collaborateurs scientifiques suivants ont pris part à ce travail: D. Yu. Panov, I. S. Mouchin, I. K. Belskaia, S. N. Rasoumovski, L. N. Korolev, G. P. Zélenkevitch.

Les expériences de la traduction ont été réalisées par la calculatrice électronique Б.Э.С.М. Cette machine a été construite par l'académicien Lébédév.

Les plus importants progrès de la technique de la dernière dizaine d'années se sont manifestés dans la construction des calculatrices électroniques sous le contrôle de programme.

Les calculatrices universelles électroniques modernes effectuent plusieurs milliers d'opérations arithmétiques dans une seconde et elles remplacent le travail de plusieurs dizaines de milliers d'hommes-calculateurs.

Il est patent que la grande vitesse avec laquelle les opérations arithmétiques s'effectuent, n'aurait aucun sens si l'on ne pouvait pas complètement automatiser le procès de calcul qui contient l'ordre établi d'exécution des opérations et qui conserve les résultats intermédiaires pour leur utilisation soivante etc.

Il faut marquer à part les commandes du passage conditionnel qui permettent à la machine, selon les résultats de calcul, de passer automatiquement de remplir l'une ou l'autre séquence des commandes, c'est-à-dire de choisir la voie des calculs suivants. On peut donc questionner la machine; elle répondra selon les circonstances «oui», ou «non».

On appelle les commandes du passage conditionnel «les points de la bifurcation de programme». Les programmes affectés à effectuer les problèmes logiques ont beaucoup de points de la bifurcation, tandis que les programmes des opérations de calcul ont peu de points pareils.

Le nombre de points de la bifurcation détermine la complexité de la structure de programme quand on le compose.

La possibilité de modifier les commandes du programme, au cours du travail, et de choisir la voie suivante de calculs selon les résultats obtenus, permet d'utiliser les calculatrices électroniques pour exécuter une vaste classe d'opérations logiques.

L'opération de la traduction automatique d'une langue en autre se réfère à la même catégorie d'opérations.

Depuis le janvier 1955, le groupe des collaborateurs scientifiques de l'Institut de Mécanique Exacte et de Technique de Calcul et les collaborateurs scientifiques de l'Institut d'Information Scientifique de l'Académie des Sciences de l'URSS commença à développer les problèmes de la traduction automatique de l'anglais en russe.

Au bout de 1955 on a reçu les premières traductions expérimentales effectuées par la calculatrice électronique БЭСМ (la construction de l'académicien Lébédév).

Ces résultats expérimentaux ont montré qu'il est possible de créer le système de l'analyse de proposition qui permet de fixer le sens de tous les mots de cette proposition en se basant sur les connaissances de la grammaire et du lexique de la langue. Le système pareil ne dépend pas pratiquement du vocabulaire et convient à la traduction non seulement du texte spécialement choisi mais il convient à tous les textes aux éléments techniques et scientifiques du profil choisi.

Il est vrai que la langue est le moyen des relations entre les gens et qu'elle représente le système déterminé dont le sens des mots et de leurs modifications peut être exprimé par les moyens lexicaux et grammaticaux. Ces moyens s'énoncent comme les règles définies du lexique et de la grammaire, comme les lois de la construction de la phrase. En conséquence, il est possible d'élaborer les règles de la traduction qui prendront en considération toutes les particularités de la langue et permettront d'établir le sens des mots et les rapports entre les mots dans la proposition. En d'autres termes, il est possible d'établir les règles de la traduction automatique qui permettront d'effectuer, à condition de leur satisfaction, la traduction sans travailler préalablement et sans réviser ensuite des phrases et des textes différents.

Il est nécessaire d'avoir le dictionnaire pendant la traduction ordinaire et pendant la traduction à l'aide de la machine.

La calculatrice électronique opère sur les nombres, c'est pourquoi les mots dans le dictionnaire doivent être représentés sous une forme de nombres, c'est-à-dire ils doivent être codés. En mettant, par exemple, les nombres du code connu Bodo (fig. 1) en conformité avec chaque lettre de l'alphabet latin et de l'alphabet russe, nous aurons la possibilité d'inscrire chaque mot anglais et sa traduction russe par le nombre définitif, qui est l'unique pour ce mot.

a — 16	f — 14	k — 19	p — 24	u — 20
b — 06	g — 10	l — 27	q — 23	v — 29
c — 22	h — 26	m — 11	r — 07	w — 13
d — 30	i — 12	n — 15	s — 05	x — 09
e — 08	j — 18	o — 28	t — 21	y — 04
				z — 25

Fig. 1. Le codage de l'alphabet latin.

Par exemple, les mots:

There, short, into, numerical, methods
 on inscrira sous l'aspect de nombres.
 2126080708, 0526280721, 12152128, 152011160712221627,
 11082126283005.

Chaque mot a son code et son numéro d'ordre (de vocabulaire) définitif.

Le vocabulaire affecté à la traduction automatique se distingue du vocabulaire ordinaire, car il a une série de renseignements supplémentaires (des indices) qui se rapportent à la grammaire du mot russe, sauf le mot russe qui correspond au mot anglais donné. Pour le nom ce sont des renseignements suivants: genre, déclinaison, thème mouillé ou dur présence ou absence du chuintant dans le thème, quel objet signifie le mot: animé ou non etc.

Pour le verbe ce sont: conjugaison, aspect etc.

Pour l'adjectif — thème mouillé ou dur, degrés de comparaison etc.

On code les principes de grammaire selon les règles établies. Ces principes occupent 39 digits binaires, se représentant une cellule de mémoire. Les codes des mots anglais et les codes de leurs indices se représentent la partie «anglaise» du vocabulaire.

La seconde partie du vocabulaire est «russe», elle comporte des mots russes inscrits selon l'ordre de leurs numéros indiqués dans la partie anglaise du vocabulaire.

On avait préparé le vocabulaire composé de 952 mots anglais et de 1073 mots russes pour les expériences, réalisés dans la machine БЭСМ. Si l'orthographe du mot du texte correspond exactement à l'orthographe du mot du vocabulaire c'est-à-dire si les deux mots sont représentés par les mêmes nombres, on peut le constater facilement à l'aide de l'opération de comparaison. Le choix des mots dans le vocabulaire est basé sur ce principe.

Pour choisir les mots qui prendront quelques terminaisons (en anglais ce sont -s, -ing, -ed, -er, -est, -e, -y, etc), il faut supprimer ces terminaisons et recommencer

les recherches du mot sans la terminaison dans le vocabulaire.

C'est très facile de trouver les mots monosémantiques dans le vocabulaire. Mais c'est plus difficile de déterminer le sens du mot dont on a besoin si le mot est polysémantique. Pour déterminer le sens du mot polysémantique qui convient à la phrase donnée il faut faire l'analyse des mots voisins, examiner leurs significations et leurs caractéristiques grammaticales. Les règles pour déterminer la signification nécessaire du mot polysémantique sont basées sur l'analyse du grand matériel et elles sont composées en schéma qu'on appelle «supplément».

Cette schéma contient de même les expressions idiomatiques. On peut ordinairement mettre en relief «le mot typique» dans chaque expression de ce genre, le mot qui se rencontre dans toutes les variantes de cette expression idiomatique.

On a donné à ce mot l'indice «polysémantique» dans le vocabulaire. Le supplément comprend le contrôle des mots voisins pour constater que le mot ne fait pas partie de l'expression idiomatique.

Si la réponse est positive, toute l'expression idiomatique se traduit de l'expression équivalente. Par exemple, le mot anglais «able» on traduit ordinairement comme l'adjectif «способный», mais en combinaison avec les formes verbales «be» (is, are, were, been, being), c'est mieux de le traduire comme le verbe «мочь».

En conséquence, avant de traduire le mot «able» en russe comme «способный», il faut examiner le mot précédent «be, are», etc. en cas de la réponse positive il faut traduire toute la locution comme «мочь».

Il faut noter que les plus simples et les plus généraux critères du choix de la signification d'un mot ou d'une expression sont les résultats du travail préliminaire du matériel analysé par les linguistes. La traduction dépend beaucoup de cas typiques qui doivent être tous prévus et pris en considération dans le programme.

Si un des mots de la phrase manque dans le vocabulaire, la mémoire conserve ce mot sans le modifier. Quand la phrase traduite sort de la machine, le mot dont on n'a pas trouvé la traduction est imprimé en lettres latines.

Le texte anglais s'introduit à la machine phrase par phrase. D'abord on enregistre ce texte sur le ruban de papier par l'intermédiaire du perforateur de lettre en conformité du code adopté Bodo.

La phrase s'introduit ensuite du ruban perforé à la mémoire intérieure de la machine.

Par exemple, la phrase codée: «There are two various numerical methods» passe dans la mémoire intérieure comme un nombre qui a soixante dix huit chiffres:

2126080708001607080021132800291607122820
05001520110807122216270011082126283005

Puis on partage la phrase en mots et commence à chercher ces mots dans le vocabulaire.

La vitesse de la traduction dépend beaucoup de recherches des mots dans le vocabulaire.

En conséquent, on doit prêter grande attention aux questions d'accélération du travail de cette partie de programme.

Voilà une de ces méthodes de recherches des mots: Les codes des mots anglais sont disposés dans le vocabulaire par l'ordre d'accroissement des nombres dont ils représentent.

En choisissant n'importe quel mot dans le vocabulaire et en comparant son code avec le code du mot cherché nous pouvons répondre à la question: où nous devons chercher le mot nécessaire dans le vocabulaire (plus haut ou plus bas). Ce procès ressemble à la méthode employée pour trouver les racines des équations selon la méthode de «position fausse».

Pour trouver n'importe quel mot dans le vocabulaire comportant 10 000 mots, suivant cette méthode, on aura besoin de 14 vérifications pareilles ($14 \approx \log_2 10\ 000$), et pour trouver le mot dans le vocabulaire comportant 30 000 mots on n'aura besoin plus que 15.

Quand le mot est trouvé dans le vocabulaire, on choisit du vocabulaire toutes les informations à propos de ce mot: le numéro de ce mot dans la partie anglaise du vocabulaire, le numéro du mot russe correspondant au mot anglais, les informations grammaticales se rapportant à ce mot russe. Cette information est une équivalent numérique du mot.

Toutes les opérations suivantes se réalisent avec ces équivalents numériques. On laisse deux cellules de mémoire pour conserver un équivalent numérique, les nombres qui désignent les parties du discours et leurs caractéristiques grammaticales on met dans la même place de cellule.

Cela permet de distinguer et de choisir automatiquement les indices nécessaires.

Les équivalents ne sont pas seulement des garages d'informations des mots. Le remplacement des mots par leurs équivalents assure l'universalité des schèmes grammaticaux et des programmes qui les réalisent. Finalement, les schèmes et les programmes deviennent indépendants de la phrase concrète et du vocabulaire.

On partage le programme de la traduction automatique en deux parties: l'analyse et la synthèse. (Fig. 2).

Le but de la première partie est de déterminer la forme grammaticale des mots russes nécessaires, leur place dans la proposition traduite, selon l'orthographe des mots anglais, leur place dans la proposition et leurs indices grammaticaux pris du vocabulaire. Les informations obtenues s'expriment par les indices et elles permettent de passer à la seconde partie du programme, c'est-à-dire, à la synthèse de la proposition russe. Les mots russes qui se trouvent dans le vocabulaire, par l'intermédiaire de ce programme, prennent la forme grammaticale correspondante aux indices obtenus par l'analyse.

La grammaire est représentée dans la phrase russe et dans la partie anglaise du vocabulaire comme les schèmes spéciaux sur les principales parties du discours: l'adjectif, l'adjectif numéral etc. L'analyse dichotomique est à la base du travail de chaque schème, c'est-à-dire c'est un système du «contrôle» de la présence ou de l'absence de n'importe quel indice grammatical morphologique ou syntaxique du mot analysé. Il est possible deux réponses sous le contrôle: la réponse positive et la réponse négative. Chacune d'eux permet de donner la conclusion définitive et de recevoir les indices nécessaires au mot en question ou de continuer le contrôle pour montrer la présence de l'indice caractéristique suivant jusqu'à ce que la réponse définitive et l'indication des indices grammaticaux qu'on doit obtenir par le mot donné ne soient pas reçues. Les parties particulières du programme sont disposées par ordre qui assure l'obtention des indices nécessaires à effectuer les actions qui suivent. Comme le but final du schème de l'analyse est d'obtenir des indices correspondants aux mots russes, les schèmes dichotomiques répondent à la question duquel mot on doit prendre ces indices. Par exemple, le nom peut prendre son cas de la préposition ou de l'adjectif numéral qui le précède.

Les schèmes de la synthèse russe ne dépendent pas de la langue dont on traduit. Les indices donnés dans le vocabulaire et recus par les programmes de schèmes

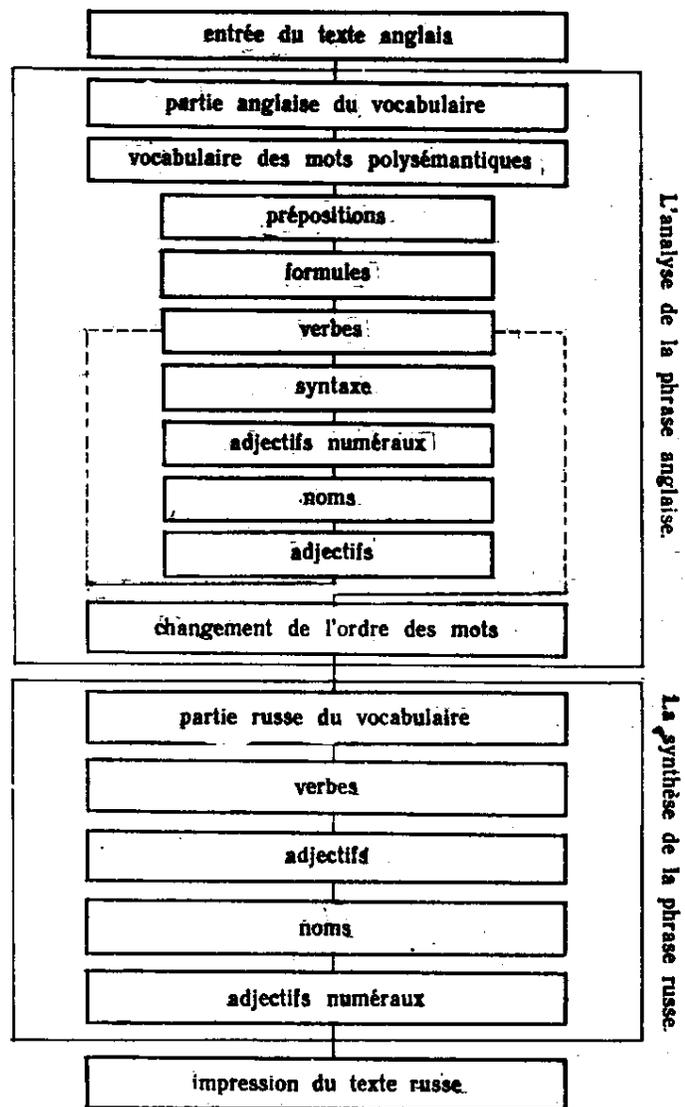


Fig. 2. Le schème du programme de la traduction automatique du russe en anglais.

de l'analyse assurent complètement l'obtention de la forme nécessaire du mot russe. De ce côté les schèmes de la synthèse russe dépendent de la partie russe du vocabulaire. La structure de la construction des schèmes russes dépend du rôle du vocabulaire. On avait admis à conserver les mots dans son essentiel, c'est-à-dire, on doit conserver les adjectifs au masculin, au singulier, au nominatif; les verbes — à l'infinitif etc. Les schèmes russes réalisent, pour la plupart, dans ce cas, le changement de terminaisons des mots pris du vocabulaire et ils tiennent compte de l'alternance de voyelles et de consonnes quand on en a besoin.

Les schèmes russes de la synthèse opèrent sur le mot et ses indices et on n'a pas besoin de prêter attention aux mots voisins. C'est la distinction principale des schèmes russes de la synthèse et les schèmes de l'analyse (des schèmes anglais dans ce cas).

Examinons les étapes successives de la traduction par la machine. Prenons, par exemple, la phrase anglaise: «The cause of this phenomenon will be considered in the following articles».

Cette phrase codée s'introduit dans la mémoire intérieure comme un nombre qui a cent quarante six chiffres

2126080022162005080028140021261205002426
081528110815281500131227270006080022281505
12300807083000121500212608001428272728
13121510001607211222270805

Puis on la repart en mots et cherche chaque mot dans le vocabulaire.

On ne trouve que cinq mots dans cette phrase dont chacun a une seule signification et les indices russes correspondants:

phenomenon — traduction: явление; le nom, neutre, I déclinaison, thème mouillé, anglais.

consider — traduction: рассматривать; le verbe, I conjugaison, aspect imparfaitif, il exige l'accusatif, angl.

article — traduction: статья; le nom, du féminin.

will — on ne le traduit pas, le verbe.

be — traduction: быть; le verbe, I conjugaison, aspect perfectif, anglais.

Il est à noter que la machine a trouvé les mots «consider» et «article» après avoir supprimé la terminaison «s»

du mot «articles» et la terminaison «ed» du mot «considered». On l'a marqué dans les équivalents de ces mots.

On marque que les autres mots ont beaucoup de significations dans le vocabulaire et on doit faire l'analyse de cette phrase pour choisir les significations nécessaires à ces mots.

Le mot anglais «in» qu'on traduit souvent comme «в» (préposition), on peut traduire dans certaines locutions autrement. Par exemple, dans la locution «interest in . . . интэрес к . . .», on le traduit comme «к» (préposition). On peut le traduire comme «с», «при», etc. On voit d'après le contrôle des mots voisins de cette phrase qu'il faut traduire «in» comme «в» (préposition).

Un autre exemple: le mot «follow». On le traduit dans la locution «follow» soit «такой» soit «следующим образом» selon certaines conditions qu'on doit accomplir. Dans les autres cas, on le traduit comme «следовать».

L'information, prise du vocabulaire, obtenue au moyen de l'analyse des mots polysémantiques s'inscrit comme la séquence d'équivalents des mots de la phrase en question. Les numéros russes d'équivalents correspondent à la traduction suivante mot à mot:

причина	этот	явление	быть
рассматривать	в	следовать	статья

Il est évident qu'on ne peut pas traduire certains mots anglais en russe (the, of, will) et on donne une remarque nécessaire à leurs équivalents («vide» dans les classes du numéro russe).

En travaillant les équivalents de mots de cette phrase, les programmes de l'analyse grammaticales complètent les équivalents initiaux par une série des indices nécessaires au travail des programmes de la synthèse russe. Par exemple, les programmes obtiennent pour le mot «article» les indices supplémentaires suivants: pluriel (parce qu'on a supprimé la terminaison «s»), puis prépositionnel (parce qu'on met devant le mot la préposition «в») Ces indices et les indices pris du vocabulaire sont suffisants pour obtenir la terminaison correcte du mot «статья» par l'intermédiaire du programme russe.

Le programme du verbe anglais obtient les indices supplémentaires suivants pour le mot «follow»: le participe présent (parce qu'on a omis la terminaison «ing») et les indices du pluriel et du prépositionnel qu'on prend du nom suivant.

A la fin du travail des programmes «anglais» de l'analyse aux équivalents et des programmes «russe» de la synthèse aux mots de sous-texte choisis dans la partie russe du vocabulaire, on recoit la traduction définitive.

Причина этого явления будет рассмотрена в следующих статьях.

Il faut mettre en relief que dès l'entrée de la phrase anglaise dans la machine, la traduction se fait automatiquement sans aucune intervention de l'homme. Les philologues et les mathématiciens ont beaucoup travaillé préalablement pour faire traduire la machine de cette façon.

Les expériences de la traduction automatique réalisée à l'U.R.S.S. et à l'étranger laissent espérer que le temps est proche où l'on réalise l'automatisation de la traduction du texte aux éléments scientifiques et techniques sur la plus vaste échelle.

Pour conclure, citons quelques exemples des phrases anglaises et leurs traductions russes, obtenues par la machine Б.Э.С.М.

When a practical problem in science or technology permits mathematical formulation, the chances are rather good that it leads to one or more differential equations. This is true certainly of the vast category of problems associated with force and motion, so that whether we want to know the future path of Jupiter in the heavens or the path of an electron in an electron microscope we resort to differential equations. The same is true for the study of phenomena in continuous media, propagation of waves, flow of heat, diffusion,

Если практическая задача в науке или технике допускает математическую формулировку, шансы довольно велики что это приводит к одному или более дифференциальным уравнениям. Это верно безусловно для обширной категории задач связанных с силой и движением, так что хотим ли мы знать будущий путь Юпитера в небесах или путь электрона в электронном микроскопе мы прибегаем к дифференциальным уравнениям. То же верно для изучения явлений в непрерывной среде, распространения волн, потока тепла, диффузии, статистического или динамического электричества, и т. д., за ис-

static or dynamic electricity, etc, except that we here deal with partial differential equations.

Equations involving more than one independent variable and the partial derivatives of the dependent variables with respect to the independent variables are called partial differential equations.

It is often impossible, however, to perform the actual elimination, and hence this transformation is of theoretical rather than practical interest.

To illustrate the use of equation 54.4 we apply it to the approximate solution of the differential equation.

There are various numerical methods for this purpose.

It is necessary to find values between which the function $f(x)$ is zero.

Suppose that both equations actually contain all the possible partial derivatives of second order.

In problems of this type numerical methods become a necessity due to absence of other methods for getting the requisite

ключением того что мы здесь будем рассматривать дифференциальные уравнения в частных производных.

Уравнения, содержащие более чем одну независимую переменную и частные производные зависимых переменных относительно независимых переменных называются дифференциальными уравнениями в частных производных.

Часто невозможно, тем не менее, выполнить действительные исключения, и следовательно это преобразование имеет теоретический скорее чем практический интерес.

Для того чтобы иллюстрировать применение уравнения 54.4 мы будем применять его для приближенного решения дифференциального уравнения.

Имеются разные численные методы для этой цели.

Необходимо найти значения, между которыми функция $f(x)$ есть нуль.

Допустим, что оба уравнения действительно содержат все возможные частные производные второго порядка.

В задачах этого типа численные методы становятся необходимостью обусловленной отсутствием других методов для получения необходимого

information out of the differential equations.

This was based on an expensive experiment done by myself and Dr. R. H. Richens, of Cambridge University, in which we worked out a method of translating small sections of selected text in foreign languages. We gave an account of this at a conference in Massachusetts in 1952, after which the International Business Machines Company, in conjunction with Georgetown University, applied our methods to give a popular demonstration which was limited to translating a few sentences from Russian into English. There is no possibility at present of translating a book as a work of art.

сведения из дифференциальных уравнений.

Это было основано на дорогом эксперименте проведенном мной и доктором R. H. Richens, от Кембриджского Университета, в котором мы разработали метод перевода малых отрывков выбранного текста на иностранные языки. Мы дали отчет о этом на конференции в Massachusetts в 1952, после которого I.B.M. компания в сотрудничестве с Джорджтаунским Университетом применили наши методы чтобы дать наглядную демонстрацию, которая была ограничена переводом нескольких предложений с русского на английский. Не имеется возможности в настоящее время перевода книги как произведения искусства.